数据故事化实践应用分析
——以数据新闻为例

■ 周霞 王萍 张韫麒 宋婧馨 陈为东
吉林大学管理学院 长春 130022

摘 要:[目的/意义]数据故事化实现数据、可视化与叙事的关联融合,有利于公众感知及认知复杂数据,提高数据重用率。[方法/过程]通过提取新华网260个数据新闻故事题目并提炼高频词,确定27个卫生健康类热点数据新闻故事。根据扎根理论并采用质性分析软件NVivo11对热点数据新闻进行三级编码,最终得到数据新闻实践的11个初始范畴,并凝练出6个主范畴。[结果/结论]得出数据故事化实践路径的理论模型及特征。数据故事可用于知识发现、数据解读、服务增值等方面。

关键词:数据故事化 数据新闻 NVivo 11 扎根理论 逻辑叙事
分类号:G253
DOI:10.13266/j.issn.0252-3116.2021.14.014

1 引言

数据科学领域致力于提升公众对数据的认知,但复杂、晦涩、枯燥、海量动态生长的数据使公众难以挖掘到有价值的数据资源,更难厘清数据之间的关联关系,导致高价值的数据资源被搁置[1]。R. Kosara和J. Mackinlay[2]指出人类擅长用故事和可视化传递信息。数据故事化(Data Storytelling)联合可视化和叙事形成语义互联,赋予数据逻辑,洞察数据隐藏价值,提升了公众对数据的认知敏锐度。数据科学领域引用数据故事化表征方式揭示数据之间的隐性关系,赋予数据发声、解释、说服和引导作用。目前,数据科学、新闻传媒、计算机科学等领域积极探索了数据的故事化,但是细粒度的实践应用路径还尚未明晰。本文从数据新闻的小视角去探索数据故事化的实践应用路径,为数据故事的制作提供理论指导。

2 相关研究

数据故事化的研究主要包括:理论框架探析、技术开发、场景应用。

(1)数据故事的理论框架。数据故事的概念在学术界尚未得到统一,相关术语有数据故事化、可视故事化、用数据讲故事、数据叙事、数据新闻等,虽然术语称谓存在差异,但都表现为“数据+可视化+叙述”的形式。朝乐门等[3]认为数据故事化是将数据关联至特定的场景,并采用叙述揭露数据关系。F. El Outa等[4]将数据故事界定为信息的结构化组织,通常会采用可视化方式表征。C. Zhang[5]认为数据故事化的理论框架包括:产品层、知识系统层、用户层,旨在实现隐性知识显性化,从而创造新知识。目前数据故事化还缺乏通用的理论模型,亟待构建指导性的理论框架[6]。

(2)数据故事的技术开发。自数据故事化提出以来,学术界致力于探索故事的自动化生成技术。当前数据故事的技术开发有:设计自动化生成系统、构架数据故事实现技术框架指南。J. Ni等[7]采用深度学习技术从灾难新闻中提取三元组并以图形的方式展示了灾难事件演化过程,发现数据叙事促进了公众深刻理解灾难事件的发展概况。朝乐门[8]采用数据科学与数据工程结合的方法,提出数据故事自动生成流程并设计故事的工程化研发参考架构。D. Shi等[9]开发了新的数据故事生成系统Calliope,它通过相关设备实现自动生成过程,将数据的电子表格转换成数据故事,并通过对照实验法验证了Calliope系统的可操作性。H.

作者简介:周霞(ORCID:0000-0002-1440-4394),博士研究生;王萍(ORCID:0000-0002-0308-7362),教授,博士生导师,通讯作者,E-mail:jlwang65@163.com;张韫麒(ORCID:0000-0002-1383-9978),博士研究生;宋婧馨(ORCID:0000-0003-2918-9268),博士研究生;陈为东(ORCID:0000-0003-4546-1305),博士研究生。
收稿日期:2020-12-11 修回日期:2021-03-25 本文起止页码:119-127 本文责任编辑:徐健

O. Obie 等[10]开发了 Gravity 系统，Gravity 系统可以整合可视化数据故事创建的全过程，并按照逻辑顺序排列可视化集合，支持连贯叙述和视觉效果，降低公众认知负荷。Y. R. Cao 等[11]提出 VisGuide 系统可以协助用户创建上下文可视化序列数据，创建结构化的数据故事。技术提高了数据故事生成的效率，但系统的普及性不高并对公众存在着操作障碍。

（3）数据故事化的应用场景。数据故事化应用场景主要有数据新闻、商务演示、学术交流[3]和开放数据[12]。许向东[13]将线型叙事、组合叙事、交互型叙事作为新闻叙事的 3 种方法，为“数据讲故事”给予了理论支持。李岩等[14]在《数据新闻手册》的基础上运用文本分析方法阐明了数据新闻讲故事的原理。V. J. C. Arevalo 等[15]认为科学交流中的故事情节能够增强科学创新。N. Ó. Brolcháin 等[16]指出数据故事化帮助了开放数据平台的用户更好地获取原始数据信息。王萍等[12]采用单案例研究并依托“5W”传播理论提出了政府开放数据故事化的实现模式。

综上所述，数据故事化研究处于摸索阶段，表现为：①相关成果主要聚焦于数据科学、新闻传播学、计算机科学等领域，在图情档学科探索较少；②故事生成的技术方法在计算机领域已经取得一定的成果，但是否能延伸到其他领域尚待验证；③需进一步探寻通用的理论框架及模型。数据故事广泛应用于数据新闻，该领域积累了丰富的数据故事案例，已相对成熟。数据故事报道方式最早可以追溯到 17 世纪，如 1821 年卫报“读者栏目”以丰富的数据图表证明了在校学生人数并非教育改革者所认为的学生数量，标志着用数据讲故事的开端[17]。随之专业的数据新闻制作平台 ProPublica Nerd Blog 、Online Journalism Blog 等[18-20]也应运而生，数据新闻的方法、手段、策略值得其他领域借鉴。本文运用扎根理论方法深度解构数据新闻的实现过程，通过小视角方式阐明数据讲故事的路径。

3 研究设计

3.1 研究方法

扎根理论是从原始资料中系统归纳出新理论，是一种灵活且系统的质性研究方法，允许建构的理论回溯到原始数据，适用于研究的探索性阶段[21-22]。数据故事化实践应用尚处于发展初期，较宜选择扎根理论作为研究方法。B. Glaser 等[23]认为“一切皆为数据”，包括访谈、观察、反思、文本、图片、视频等，可用于解决社会科学中理论与经验研究的脱节问题。由于数据新闻呈现的结果不是原始资料，故本文不分析数据新闻的内容，只观察、归纳数据新闻的文本、数据和图片的表现特质，其本身特征仍属于原始资料范畴。本文根据扎根理论思想：问题产生→数据收集→数据处理→理论初构[24]，得出整体研究思路如图 1 所示：

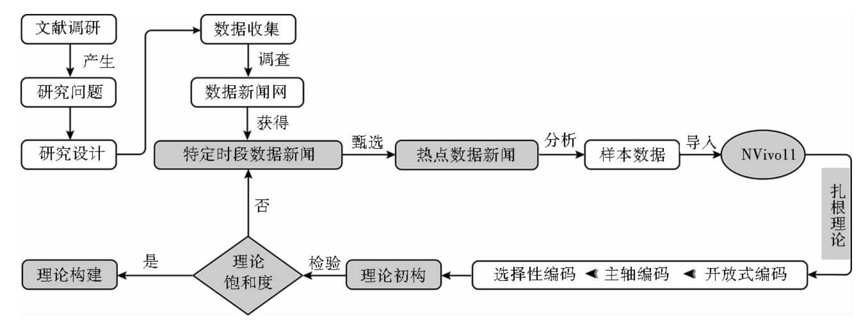


图 1 整体研究思路

如图 1 所示：①通过文献调研提出研究问题，即实践应用过程中数据新闻故事是如何生成的？②搜集新华网特定时间段内的数据新闻，对新闻题目中出现的高频词，用词云确定热点数据新闻，并作为样本；③将样本数据新闻导入质性研究工具 NVivo11 进行开放式编码，形成初始概念；④将初始概念归纳出初始范畴并关联，形成主轴编码；⑤在主轴编码基础上开展选择性编码，建立主范畴之间的逻辑关系；⑥另选取样本总量 1/4 的数据新闻作为理论饱和度检验，理论饱和后则构建数据新闻实践路径的理论模型。

3.2 样本选取

新华网是国内较早开展数据新闻实践的媒体机构，其《数据新闻》栏目的稿件覆盖范围广泛且专业程度居于同类媒体领先地位，已累计播发近千期[25]。新华网《数据新闻》栏目分类有 10 种，由于“政经事”与“讲习所”为政策解读类新闻，本文不统计该类栏目下的新闻数量，选取其余 8 种类型，经过筛选得到有效数据新闻 260 个（搜集时间为 2019 年 7 月 9 日至 2020 年 7 月 9 日），新闻类型及数量分布如表 1 所示：

周霞，王萍，张韫麒，等. 数据故事化实践应用分析——以数据新闻为例[J]. 图书情报工作,2021,65(14):119－127.

表 1　新华网数据新闻类型和数量分布

类型	数量(个)
数据观	8
第一时间	48
数问民生	128
数极客	3
涨知识	18
人文说	9
健康解码	31
漫生活	15
总计	260

新华网数据新闻内容涉及生活的多个方面,覆盖领域较广,重点关注民生及热点事件,其中"数据观"主要展示数据可视化技术;"第一时间"是以数据故事化形式讲述国内大事新闻;"数问民生"发布百姓关注的生活事件;"数极客"是黑科技展示,发布内容较少;"涨知识"是普及日常生活的基本知识;"人文说"借助数据分析人物生命历程轨迹和文化;"健康解码"公布健康类的数据新闻;"漫生活"用数据图文展示动漫。

数据新闻题名可以映射主题内容,分析题名有助于识别出热点选题,采用 BDP 生成词云如图 2 所示:



图 2　新华网数据新闻题名词云

260 个数据新闻主要涵盖:确诊、北京、新增、病历、肺炎等词,青睐于卫生健康类选题。由于新冠疫情的原因,卫生健康成为社会热切关注的焦点,本文最终选定 27 个卫生健康类数据新闻为扎根样本。

3.3　数据编码过程

数据新闻是数据故事化的应用之一,其策划、选题、主题提炼、数据萃取及可视化等共同构造了数据故事[26]。本研究数据编码过程由 2 位博士生完成,编码结果由 4 位博士生及 1 位领域专家共同商讨,再由 1 位博士生进行理论饱和度检验。

3.3.1　开放式编码

开放式编码是扎根理论的初始阶段,是将收集的资料打碎并赋予概念的过程[27]。新华网数据新闻导出格式为 JPG 和 PNG,属于图片格式,每张图片代表一个数据新闻故事,由叙事文本、数据、可视化图表等内容组成。本文借助质性研究工具 NVivo11 对新闻故事逐个编码,归纳其中有关数据故事化构造的表征特质,每个数据新闻出现一个特质或多个相同特质仅做一次编码。通过不断地比较、探讨及梳理,共得到 413 条原始语句,50 个初始概念(a1-a50),经过合并提炼出 11 个初始范畴(A1-A11)。由于一句话、一段文字、一个图表等表现的特质作为开放式编码资料的表征形式,囿于篇幅,仅列举了部分资料,开放式编码结果见表 2。

3.3.2　主轴编码

主轴编码是在开放式编码的基础上,通过聚类发现范畴之间的关系[28]。基于开放式编码的 11 个初始范畴,进行关联、总结、整合,最终得到 6 个主范畴(B1-B6):多源数据、叙事声音、视觉效果、叙述逻辑、关联应用、辅助信息,主轴编码结果见表 3。

3.3.3　选择性编码

选择性编码是深化主范畴的关联关系、凝练核心范畴和用"故事线"描述现象的重要阶段[29,30]。本文对主轴编码形成的 6 个主范畴系统分析,梳理主范畴与数据故事化实践应用的关联关系,从而建立主范畴的关系结构见表 4。

3.3.4　理论饱和度检验

理论饱和度是对扎根理论应用后最终结果的检验,主要运用额外数据资料检验是否出现新范畴,从而决定是否停止采样[31],选取新华网 7 个卫生健康类数据新闻检验理论饱和度。由 1 位博士生进行饱和度检验,未发现数据新闻的新范畴,因而认为该样本达到了饱和。

4　数据故事化实践路径的理论模型构建

4.1　数据新闻原有生产过程

M. Lorenz[32]给出了数据新闻的生产过程:数据－过滤－可视化－故事,见图 3。

数据经过过滤和可视化并在特定情境下进行阐释,从而形成故事。在此过程中,原有数据会转化为有意义的知识,增加了公众的感知价值,特别是当复杂的数据(事实)被归结为一个清晰的故事时,更易于公众理解和认知。M. Lorenz 的数据新闻生产过程的研究对数据故事化研究具有宏观的理论指导意义。本文在 M. Lorenz 的数据新闻生产过程的理念上,更加关注叙事者的主导地位及公众对数据故事生成路径的微观认知,因而需要进一步细粒度地构建数据故事化的实践路径。

表 2 卫生健康类数据新闻表征特质的开放式编码结果及范畴

初始范畴	初始概念	原始资料(部分)
A1 故事题名	a9 疑问类题目	《湖北一夜之间新增新冠肺炎病例 14840 例,为什么?》
	a13 反问类题目	《来自各种理由的睡不着? 你的睡眠还好吗?》
	a14 引导类题目	《这里有一份属于你的身体健康说明书》
	a22 概括类题目	《北京 12 名确诊病例主要活动轨迹》
	a23 呼吁类题目	《别再喝红糖水了,教你科学治疗缺铁性贫血》
	a25 趣味类题目	《太"秃"然? 防脱发小妙招了解一下!》
A2 数据来源	a6 数据来源高校	《美国约翰斯・霍普金斯大学》
	a20 数据来源商业领域	《中国移动互联网数据库 2020 年 2 月》
	a32 数据来源政府	《中国疾病预防控制中心》
	a42 数据来源公共领域	丁香园《2019 年国民健康洞察报告》
A3 可视化统计图表	a1 坐标图 a3 散点图 a5 矩阵图 a7 数据地图 a11 等级图表 a15 时间轴 a18 折线图 a19 环状图 a21 柱状图 a26 饼图 a29 条形图 a46 动态图表	数据新闻里使用了坐标图、散点图、矩阵图、数据地图……(由于可视化统计图表均使用图形,故本文原始资料以文字代替)
A4 图画展示	a37 漫画形式 a38 简笔画	数据故事里使用了漫画及简笔画(由于漫画形式、简笔画均使用图画,故本文原始资料以文字代替)
A5 单数据凸显	a41 数据凸显	数据新闻单数据凸显采用字体变大或改变数字颜色……
A6 故事背景	a43 开场白	冠心病、高血压、呼吸系统疾病……吸烟严重危害人民的身心健康。5 月 31 日是世界无烟日,让我们一起向烟草说"不"。
A7 叙述句法	a2 解释型叙述形式	艾滋病离你我并不遥远? 截止 2018 年 8 月底,全国报告存活感染者 85 万,死亡 26.2 万例……;
	a4 启发式叙述形式	您是糖尿病高危人群吗? 具备以下因素之一,即为糖尿病高危人群(1)超重与肥胖;体重指数≥ $28kg/m^2$;(2)高血压,收缩压≥140mmHg 和(或)舒张压≥90mmHg……
	a8 呼吁类叙述形式	预防乳腺癌你要这么做! 坚持锻炼、避免熬夜、坚持母乳喂养、坚持健康体检、健康饮食、戒烟戒酒……
	a24 疑问类叙述形式	你吃的健康吗? 你是运动达人吗? 你的眼睛清晰吗? ……
	a33 问题-自答叙述形式	得了缺铁性贫血咋办? 药物治疗、病因治疗、输血治疗……
	a35 对比类叙述形式	2019-2020 年不同时段移动互联网用户概况:(图表对比)2019 年春节 8.39 亿、2020 年平日 8.45 亿、2020 年春节 8.50 亿、2020 年春节后 8.49 亿. . . .
	a45 陈述型叙述形式	湖北省临床诊断病例:是尚不具备病原学证据,但具备临床表现、肺炎影像学特征的患者……
A8 故事制作者	a12 策划	策划:刘舒
	a17 文案	文案:蒋燕、邓思敏(实习)
	a27 监制	监制:展鹏 彭嘉靖
	a28 编辑	编辑:刘品彤
	a34 设计	设计:冯光明
A9 外链应用	a30 数据新闻客户端二维码 a31 数据新闻微信公众号 a40 数据新闻二维码	数据新闻给出了二维码或公众号(由于二维码和微信公众号均以图标呈现,故本文原始资料以文字代替)
A10 其他信息	a10 设计制作单位	设计制作:陕西频道
	a16 备注信息	备注信息,时段定义:2019 年春节指 2019 年 2 月 4 日-10 日,2020 年平日指 2020 年 1 月 2 日-8 日,2020 年春节指 2020 年 1 月 24 日-2 月 2 日 2020 年春节后指 2020 年 2 月 3 日-9 日
	a36 联系方式	联系电话:010-88050306
	a39 版权声明	新华网版权作品,请获取授权后转载
	a44 网址链接	更多精彩尽在 http://news.cn/datanews
A11 叙事形式	a47 类比叙事	我国目前口罩和呼吸防护相关企业超 8.5 万家,新增超 3.1 万家,与去年同期比较,增速高达 2013.54%……
	a48 网状叙事	展开平铺报道:今年"五一"连休 5 天,是大小长假政策实施多年来的首次。宅了几个月的人们,面对着大好春光和 5 天假期,是选择继续宅着还是出行呢?
	a49 线性叙事	时间轴:2 月 5 日晚 10 点(位于武汉会展中心的江汉方舱医院率先启用)→2 月 11 日(黄陂区体育馆方舱医院启用)→2 月 12 日……
	a50 延伸叙事	通过对 2013 年和 2018 年全球护士数据对比,预测 2030 年前,护士毕业生总数平均每年需增 8%,才可解决所有国家的短缺……

表 3　卫生健康类数据新闻表征特质的主轴编码结果

主范畴	初始范畴	范畴内涵
B1 多源数据	A2 数据来源	提供原始数据来源,便于受众再次创作
B2 叙事声音	A8 故事制作者	参与故事制作的主要人员
B3 视觉效果	A3 可视化统计图表 A4 图画展示 A5 单数据凸显	利用可视化给受众带来感知上的美好体验
B4 叙事逻辑	A1 故事题名 A6 故事背景 A7 叙事句法 A11 叙事形式	借助叙事使数据讲故事符合受众的认知
B5 关联应用	A9 外链应用	提供多设备的跨屏交互
B6 辅助信息	A10 其他信息	提供公众了解故事的相关信息

表 4　主范畴与数据故事化实践的关系

典型关系	关系结构
多源数据→数据故事化实践	数据来源对数据新闻实践起核心作用
叙事声音→数据故事化实践	叙事人员对数据新闻实践起主导作用
视觉效果→数据故事化实践	视觉效果对数据新闻实践起支撑作用
叙事逻辑→数据故事化实践	叙述逻辑对数据新闻实践起推动作用
关联应用→数据故事化实践	关联应用对数据新闻实践起扩散作用
辅助信息→数据故事化实践	辅助信息对数据新闻实践起协助信息

Data　Filter　Visualize　Story　rising value to public

图 3　数据新闻的生产过程[32]

4.2　理论模型构建

基于前文对数据新闻的编码,数据新闻的可视化操作工具及数据处理过程无法通过编码获取,为了保证数据新闻实践路径的完整系统性,通过文献调查补充了数据可视化转化需要的隐藏信息[3,33-35]。最终从叙事声音、多源数据、视觉效果、叙事逻辑、关联应用、辅助信息 6 个维度构建了数据故事化实践路径的理论模型,如图 4 所示:

4.3　理论模型分析与讨论

4.3.1　叙事声音

叙事声音即叙事者,包括策划、编辑、监制、设计等人员,数据可视化过程中还需要技术人员(IT 技术员、数据分析师等)。数据故事的叙事声音可以是人或者智能设备,打破了传统叙事中"作者"的叙事话语权。叙事声音主导着数据故事的制作,贯穿了故事"选题-

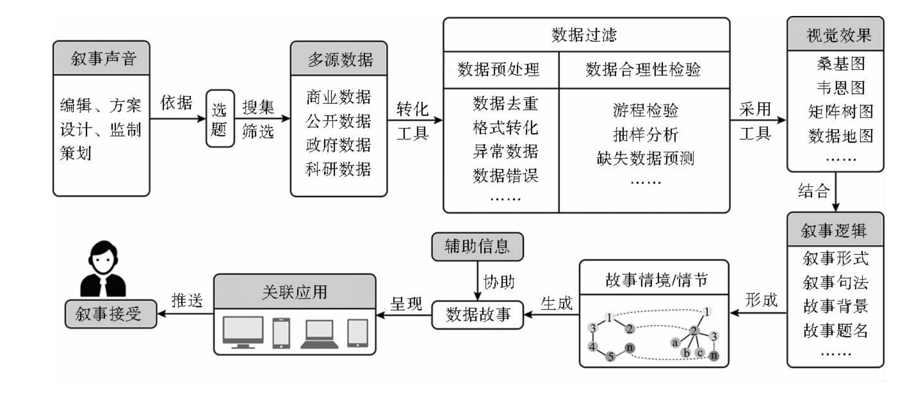


图 4　数据故事化实践路径的理论模型

数据过滤-视觉效果-叙事"的始终。叙事者专业水平直接影响数据故事的成品效果,因而叙事声音的配备需要依据数据故事的选题。在 27 个卫生健康类数据新闻中有技术人员借用 Excel、PowerBI、Tableau、Python、R 语言等工具对数据进行预处理、合理性检验、视觉效果实现等,从而构成了分工明确的叙事声音链。

4.3.2　多源数据

数据是数据故事化的核心,主要来源于政府开放数据、公共数据、科研数据和商业数据等,如《图说 90 后的养生焦虑》的数据来源有 3 个:丁香园《2019 国民健康洞察报告》《阿里数据》、Analysys 易观《90 后健康养生大揭秘 2019》,说明了数据来源多样性及客观严谨性。故事选题分为主题驱动(调查型数据新闻)和数据驱动(常规型数据新闻)两类[36],其中主题驱动围绕故事主题采集数据,选题偏重于社会民生等热点事件;数据驱动是基于已有数据集来确定选题。主题驱动较数据驱动更能体现数据新闻传播的实效性,故而备受青睐[37]。新华网 27 个数据新闻主要为

主题驱动类型,更符合受众偏好,传播效率高。确定数据后,还需:①数据预处理。叙事者需要进行数据去重、格式统一以及剔除异常和错误数据等工作,以保证数据准确性。②数据合理性检验。数据合理性检验是对数据预处理结果的再次验证,确保无异常值及非理性数据。

4.3.3 视觉效果

根据图片优势效应[38],数据新闻的视觉效果不仅提高公众的外显记忆,更能刺激其内隐记忆,从而调节公众认知的意义建构,降低数据解读的复杂度。视觉效果由可视化统计图表、图画展示、单数据凸显3个范畴构成,其中可视化统计图表是关键范畴。可视化是以数字化图像形式展现,运用计算机图像学、图像处理技术将数据转换成图形,直观揭示数据中隐含的信息[39]。新华网27个卫生健康类的数据新闻的可视化统计图表和图画展示类型丰富,如动态图表常与时间轴、数据地图结合,以动态的方式表征复杂的数据,从而更好地协助理解;单数据凸显是指在文本中以醒目颜色和字体大小显现单个数字,有利于公众快捷的捕捉关键信息。数据新闻的视觉效果本质上是对新闻故事删繁就简,强调主题,将清晰、简洁的故事主旨传递给公众。

4.3.4 叙事逻辑

数据新闻的叙事逻辑是通过叙事形式、叙述句法、故事题名、故事背景4个范畴描述故事发生的全景,从而构成完整的叙事结构。数据新闻的叙事形式表现为事件间的连接规律,主要的叙事形式有:①线性叙事。通常按照事件演进过程,采用时间轴表现故事的开端、发展、高潮、结局。②延伸叙事。延伸叙事是对叙事主线进行延伸,形成联想叙事、预测分析及时空拓展。③类比叙事。类比叙事是根据"变量"列举不同类别信息以发现数据规律。④网状叙事。网状叙事是通过信息图展开平铺方式报道[40]。本研究中的27个数据新闻从传统单一节点转向点面结合的叙述形式,多种叙事形式通常交叉使用,如《来了!新型肺炎疫情最新事实》同时采用线性、网状、类比3种叙事形式。叙述句法有解释型、启发式、呼吁式、疑问式等,丰富了数据的表现形态。故事题名是故事内容的高度提炼,而故事背景具有引导性,有利于叙事者营造氛围。叙事话语灵活多变,叙事逻辑有利于连贯多个故事情节(情节1+情节2+……情节n)并关联匹配到多种情境(情境1+情境2+……情境n),从而实现故事的逻辑顺应。

4.3.5 关联应用

数据新闻的关联场景是受众在何种载体下接受数据新闻故事,载体包括手机端、平板端、网页端。27个卫生健康类数据新闻虽然设置了微信公众号、手机端、PC端等接受场景,但存在以下问题:数据新闻应用到不同的设备环境下未能差异化适配;数据新闻也未设置反馈环节,受众无法评价数据新闻质量的优劣。

4.3.6 辅助信息

辅助信息是为受众提供更多了解数据新闻的相关信息,包括设计制作单位 、备注信息、联系方式、版权声明、网址链接等。辅助信息是数据故事化整个实践过程中的外延信息,为受众进一步获取故事信息提供相关渠道,有利于受众持续性的关注数据故事的发展。

4.4 数据故事化的特征

根据数据故事化的实践路径归纳出3个特征:①利用故事背景引导受众感知情境。数据故事通过引入关联性、丰富性、趣味性的背景导读信息,激发受众的好奇心理和情境依赖。受众在特定的故事情境下将明确角色设置、语义及行为[41]。②融入信息图表刺激受众的视觉。根据数据类型选择适配的图表,不同的信息图表导致数据应用场景也随之变化[42],K线图适合股票交易类数据、词云适合展现大量文本和文本主题的视觉效果,信息图表刺激着受众的视觉感知。③使用逻辑故事线联结受众认知图式。故事线并非泛指故事情景线索,也称故事情节或知识相关性(即知识链),强调事物之间的逻辑关系或思维结构[43]。受众对故事的理解受限于多种要素,创作者在深化故事主题和结构时需注重受众的逻辑感知体验。故事线作为数据故事的主要脉络,其设计的核心思想是使受众深层次参与到数据故事中,引起高阶思维的分析、评价、创造,实现知识的隐性到显性嬗变。数据故事可以借鉴艺术、展览行业的故事线方法,层层铺展故事情节,逐步引导受众理解[44-45],丰富认知图式。

5 数据故事化的实践应用

数据存在于政府、商业、科研等多个领域,数据故事在多个学科取得研究进展并助推领域知识的发现和传播。如何应用数据及说何种故事是数据故事化的意义[46]。本文探究数据故事化的实践应用,有助于促进其纵向延伸和横向拓展。

5.1 用于知识发现

牛力等[47]将数字人文的理念、方法和技术应用于档案领域,提出了档案数据的"发现-重构-故事化"

为主线的研究路径,重构档案知识,多视角叙述档案故事,发掘档案数据价值。例如,图博档馆藏资源卷帙浩繁,拥有海量数据。基于这些特点,首先,对多源异构数据资源进行深层次序化、细粒度的知识组织,实现不同维度的数据资源向知识资源转化;其次,依据数据故事化原理,将已经实现知识转化的数据关联,并采用叙事学方法(如空间叙事、线性叙事、网状叙事)完整地表达数据、知识、情景与事件之间清晰的层级联系,形成由点到面的情节层次结构;最后,根据用户的需求,选择合适的可视化呈现方式。将数据故事应用于知识发现,有利于公众更好地感知、解读和认知复杂数据,从而使受众更好地理解和洞悉数据蕴涵的内在知识与关联语义。

5.2 用于数据解读

大数据技术能够挖掘数据并呈现可视化结果,但缺少叙事情节,不利于公众解读数据。数据故事揭示了数据隐在的逻辑关系,通过叙事使海量数据具备了可读性,帮助受众把控数据。例如,在商业领域可以开发数据故事应用系统或 APP,当公众把数据输入到系统或 APP 时,会自动生成数据故事,故事降低了公众数据认知负荷。另外,企业通常采用故事方式作为品牌营销手段,但是数据时代,企业开始关注消费者为主体的故事,以此拉近与消费者的心理距离,有利于形塑品牌和产品的形象,进而提高消费者的参与意识[48]。企业根据运营目标,采集并过滤用户生成内容,在特定场景制作出符合消费者认知的数据故事,形成与消费者记忆相匹配的事件,从而使消费者在认识产品过程中,快速了解产品,做出合理决策。

5.3 用于服务增值

数据故事能够根据用户交互行为提供个性化服务,实现服务增值[49]。例如,互联网促进了"读屏时代"的发展,培养了读者"碎片化、短平快"的阅读方式,数据故事化的表征形式高效地迎合了读者阅读习惯[34]。在数据故事中,信息图表适合叙事空间结构、位置、细节的表达,而文本则适合抽象信息、口头概念及逻辑条件的传递[50],信息图表与文本联合编码能同时激活读者的多种认知图式类型,提高读者阅读效果。数据故事通过线性、网状、类比、组合等多种叙事形式生成多线程、模块化的叙事模块,读者能够根据自身需求选择单一模块或多种模块得出阅读结果。图书馆可以采用读者阅读的行为数据生成普适性的故事,通过网络平台、社交系统、应用 APP 等新媒体工具展示和宣传,有利于读者选择阅读书籍,从而推动图书馆的服务增值。

6 结语

数据故事化能够帮助用户更好地理解、解读和重用数据及挖掘数据潜在价值,具有广阔的应用前景。数据故事化与数据新闻具有相同的元素和叙事手法,后者作为前者的主要应用场景。本文通过扎根 27 个卫生健康类数据新闻,得到数据故事化实践路径的理论模型及特征。然而,数据故事化目前处于探索阶段,仍存在诸多障碍,需要多学科领域共同丰富其理论及实现方法。数据故事化可应用于知识发现、数据解读、服务增值方面。后续将扩大数据故事化的研究样本并探究相关领域的具体应用,如政府开放数据或科研数据的故事化,以提升此类数据的可理解性、逻辑性、灵活性、趣味性和可读性,便于公众重用数据。

参考文献:

[1] 史雅莉,赵童,杨思洛. 引证视角下科研用户的数据认知行为研究——基于扎根理论方法[J]. 情报理论与实践,2020,43(6):49－55.

[2] KOSARA R, MACKINLAY J. Storytelling: The next step for visualization[J]. Computer, 2013,46(5): 44－50.

[3] 朝乐门,张晨. 数据故事化:从数据感知到数据认知[J]. 中国图书馆学报,2019,45(5):61－78.

[4] El OUTA F, FRANCIA M, MARCEL P, et al. Towards a conceptual model for data narratives[C]//International conference on conceptual modeling. Berlin: Springer, 2020:261－270.

[5] ZHANG C. An open data storytelling framework for organizational knowledge management[C]//17th International conference on intellectual capital, knowledge management & organisational learning ICICKM 2020. UK: Academic conferences international limited reading, 2020: 415－420.

[6] 易旎,朝乐门,张晨. 可视故事化:特征、方法与应用[J]. 情报资料工作,2021,42(2):63－72.

[7] NI J, LIU X, ZHOU Q, et al. A knowledge graph based disaster storyline generation framework[C]//2019 Chinese control and decision conference. Piscataway: IEEE, 2019:4432－4437.

[8] 朝乐门. 数据故事的自动生成与工程化研发[J]. 情报资料工作,2021,42(2):53－62.

[9] SHI D, XU X, SUN F, et al. Calliope: Automatic visual data story generation from a spreadsheet[J]. IEEE transactions on visualization and computer graphics,2020,27(2):453－463.

[10] OBIE H O, CHUA C, AVAZPOUR I, et al. Authoring logically sequenced visual data stories with gravity[J]. Journal of computer languages, 2020,58(1): 1－13.

[11] CAO Y R, PAN J Y, LIN W C. User-oriented generation of contextual visualization sequences[C]//Extended abstracts of the 2020 CHI conference on human factors in computing systems. New York: Association for computing machinery,2020: 1－8.

[12] 王萍,周霞. 政府开放数据故事化实现模式研究[J]. 文献与数据学报,2020,2(2): 3 – 12.
[13] 许向东. 转向、解构与重构:数据新闻可视化叙事研究[J]. 国际新闻界,2019,41(11):142 – 155.
[14] 李岩,李赛可. 数据新闻:“讲一个好故事”? ——数据新闻对传统新闻的继承与变革[J]. 浙江大学学报(人文社会科学版),2015,45(6):106 – 128.
[15] AREVALO V J C,VERBRUGGE L N H,SOOLS A, et al. Storylines for practice: a visual storytelling approach to strengthen the science-practice interface[J]. Sustainability science,2020,15(4): 1 – 20.
[16] BROLCHĀIN N Ó,PORWOL L,OJO A, et al. Extending open data platforms with storytelling features[C]//Proceedings of the 18th annual international conference on digital government research. New York:Association for computing machinery,2017:48 – 53.
[17] 刘银娣. 从海量数据中挖掘新闻故事:数据新闻内涵探析[J]. 中国出版,2016(9):30 – 33.
[18] ProPublica nerd blog[EB/OL]. [2020 – 09 – 04]. https://www.propublica. org/nerds/.
[19] Online journalism blog[EB/OL]. [2020 – 09 – 04]. https://onlinejournalismusblog. com/.
[20] 张馨月. 新华网数据新闻故事化叙事研究[D]. 海口:海南师范大学,2019.
[21] CORBINJ M, STRAUSS A L. 质性研究的基础 – 形成扎根理论的程序与方法[M]. 朱光明,译. 重庆:重庆大学出版社,2015:1.
[22] 吴毅,吴刚,马颂歌. 扎根理论的起源、流派与应用方法述评——基于工作场所学习的案例分析[J]. 远程教育杂志,2016,35(3):32 – 41.
[23] GLASER B, STRAUSS A L. The discovery of grounded theory: Strategies for qualitative research[M]. Chicago:Aldine Pub. Co. , 1967:4 – 6.
[24] 胡媛,艾文华,胡子祎,等. 高校科研人员数据需求管理影响因素框架研究[J]. 中国图书馆学报,2019,45(4):104 – 121.
[25] 新华网《数据新闻》栏目[EB/OL]. [2020 – 12 – 13]. http://www. xinhuanet. com//politics/2015-11/02/c_1117015668. html.
[26] 欧阳云玲,郭馨泽,张庆庆. 新华网数据新闻框架构建分析[J]. 中国出版,2019(22):49 – 52.
[27] 朱馨叶,张小倩,李桂华. 图书馆阅读推广活动激励机制研究——基于 2018 年国内图书馆“世界读书日”活动案例[J]. 大学图书馆学报,2019,37(4):71 – 78.
[28] 胡蓉,赵宇翔,朱庆华. 移动互联环境下用户跨屏行为整合分析框架——基于扎根理论的探索[J]. 中国图书馆学报,2017,43(6):113 – 129.
[29] 王瑞花. 基于扎根理论员工知识囤积问题研究[J]. 情报理论与实践,2018,41(11):107 – 112,78.
[30] 韦草原,王健,张贵兰,等. 基于扎根理论的科学数据用户感知价值概念模型研究[J]. 情报杂志,2018,37(5):182 – 188.
[31] 杨峰,姚乐野. 危险化学品事故情报资源的情景要素提取研究[J]. 情报学报,2019,38(6):586 – 594.
[32] LORENZ M. Data driven journalism: What is there to learn[C]//IJ-7 innovation journalism conference, Brussels:European journalism centre,2010: 8 – 18.
[33] 刘英华. 数据新闻实战[M]. 北京:电子工业出版社,2016.
[34] 方浩,尚媛媛,张锐,等. 数据新闻中信息图表的阅读效果:来自眼动的证据[J]. 图书情报工作, 2019,63(8):74 – 86.
[35] 唐雪杰. 数据新闻的叙事研究[D]. 保定:河北大学,2019.
[36] 方洁,高璐. 数据新闻:一个亟待确立专业规范的领域——基于国内五个数据新闻栏目的定量研究[J]. 国际新闻界,2015,37(12):105 – 124.
[37] 赵杰. 我国四大门户网站数据新闻发展现状探析[J]. 新闻研究导刊,2020,11(1):197 – 198.
[38] 陈为东,杨雪. 基于图片优势效应的移情式教学设计[J]. 现代教育技术,2014,24(11):23 – 29.
[39] 白丹卉. 数据新闻中新闻叙事模式的变革——以《新京报》数据新闻为例[J]. 新闻世界,2020(3):43 – 46.
[40] 孟笛. 数据新闻生产特征及叙事模式——基于数据新闻奖提名作品的实证研究[J]. 当代传播,2016(6):23 – 26.
[41] 李旭晖,吴燕秋,王晓光. 基于角色关联的叙事型文化遗产知识表示方法[J]. 图书情报工作, 2017,61(9):116 – 122.
[42] 蚂蚁金服 – 让数据栩栩如生[EB/OL]. [2020 – 11 – 29]. https://graphin. antv. vision/zh.
[43] 廖红. 对“主题展开方式之故事线和知识链”的认识[C]//科技馆研究文选(2006 – 2015). 北京:中国科学技术馆,2016:297 – 303.
[44] TANG T, RUBAB S, LAI J, et al. iStoryline:Effective convergence to hand-drawn storylines[J]. IEEE transactions on visualization and computer graphics,2018,25(1):769 – 778.
[45] Storyline online [EB/OL]. [2020 – 12 – 11]. https://www. storylineonline. net/.
[46] 崔冰,郑昕遥. 国内数据新闻的呈现与表达——以网易数读为例[J]. 新媒体研究,2019,5(14):10 – 12.
[47] 牛力,高晨翔,张宇锋,等. 发现、重构与故事化:数字人文视角下档案研究的路径与方法[J]. 中国图书馆学报,2021,47(1): 88 – 107.
[48] 徐岚,赵爽爽,崔楠,等. 故事设计模式对消费者品牌态度的影响[J]. 管理世界,2020,36(10):76 – 95.
[49] 张晨. 基于用户交互行为分析的数据故事建模方法研究[J]. 情报理论与实践,2021,44(2):185 – 191.
[50] 王海智. 可视化叙事在数据新闻中的应用研究[J]. 传媒,2021(1):39 – 41.

作者贡献说明:

周霞:论文撰写、修改及数据搜集分析;
王萍:负责论文的构思及框架指导;
张韫麒:论文修改及校稿;
宋婧馨:论文修改及数据分析;
陈为东:论文修改、图文润色及数据分析。

周霞，王萍，张韫麒，等. 数据故事化实践应用分析——以数据新闻为例[J]. 图书情报工作,2021,65(14):119－127.

Practical Application Analysis of Data Storytelling: Taking Data News as an Example

Zhou Xia　Wang Ping　Zhang Yunqi　Song Jingxin　Chen Weidong
School of Management, Jilin University, Changchun 130022

Abstract: [Purpose/significance] Data storytelling realizes the association and fusion of data, and narration, which is conducive to the public's perception and cognition of complex data, and improves the data reuse rate. [Method/process] The author extracted 260 data news story titles from Xinhuanet and refined high-frequency words, based on this, identified 27 health-related hot data news stories. Based on the grounded theory and using the qualitative analysis software NVivo11, the author carried out three-level coding. Finally, 11 initial categories and 6 main categories of data news practice were obtained. [Result/conclusion] Get the theoretical model and characteristics of the practice path of data storytelling. Data stories can be used for knowledge discovery, data interpretation, service value-added and other aspects.

Keywords: data storytelling　data news　NVivo 11　grounded theory　logical narration